(12)

12

HOW MANY IID SAMPLES DOES IT TAKE TO SEE ALL THE BALLS IN A BOX?

Thomas M. Sellke

TECHNICAL REPORT No. 460
OCTOBER 26, 1992

Prepared Under Contract
N00014-92-J-1264 (NR-042-267)
FOR THE OFFICE OF NAVAL RESEARCH

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305-4065

HOW MANY IID SAMPLES DOES IT TAKE TO SEE ALL THE BALLS IN A BOX?

Thomas M. Sellke

TECHNICAL REPORT No. 460
OCTOBER 26, 1992

Prepared Under Contract
N00014-92-J-1264 (NR-042-267)
FOR THE OFFICE OF NAVAL RESEARCH

Professor Herbert Solomon, Project Director

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305-4065

How Many IID Samples Does it Take to See All the Balls in a Box?

Thomas M. Sellke
Purdue University

Abstract

Suppose a box contains m balls, numbered from 1 to m. A random number of balls are drawn from the box, their numbers are noted, and the balls are then returned to the box. This is done repeatedly, with the sample sizes being iid. Let X be the number of samples needed to see all the balls. This paper derives a simple but typically very accurate approximation for EX in terms of the sample size distribution. The justification of the approximation formula uses Wald's identity and Markov-chain coupling.

1. Introduction

Suppose we have a box containing m identical white balls. Let $K_1, K_2, \ldots$ be iid random variables taking positive integer values. We randomly sample $K_1 \wedge m$ balls without replacement, paint the sampled balls red, and return them to the box. Then $K_2 \wedge m$ balls are sampled, painted red, and returned to the box, etc. Let X be the number of samples needed to paint all the balls red. When $\max\limits_{1 \leq i \leq X} K_i$ is with high probability small compared to m, a good approximation for EX is given by

$$(1.1) \qquad \frac{\sum\limits_{i=1}^{m} \frac{1}{i}}{\sum\limits_{i=0}^{m-1} \frac{1}{m-i} P\{K > i\}} + \frac{\sum\limits_{r=1}^{m-1} \frac{1}{m-r} P\{K > r\} \sum\limits_{j=1}^{r} \frac{1}{m-j+1}}{\left[\sum\limits_{i=0}^{m-1} \frac{1}{m-i} P\{K > i\}\right]^2}.$$

(A K without a subscript represents a generic K_i.) For instance, if $m = 10$ and the $(K_i - 1)$'s are binomial $(4, \frac{1}{2})$, then the true value of EX is 8.8937, while (1.1) gives 8.8933. For $m = 20$ and $(K_i - 1)$'s which are binomial $(4, \frac{1}{2})$, the values are $EX =$ 22.90753529760067 and (1.1) = 22.90753529760074. For $m = 10$ and K_i's which are uniformly distributed on $\{1, 2, 3, 4, 5\}$, $EX = 8.74239$ and (1.1) = 8.74236. For $m = 20$, the values are $EX = 22.740208948996$ and (1.1) = 22.740208948981. (The true values were computed by Jacek Dmochowski using exact recursive formulas suggested by Larry

Shepp.) It is difficult to determine the size of the approximation error when m is much larger than 20. Even with double-precision computation, the round-off error seems to dominate the true approximation error.

The justification of formula (1.1) involves Wald's identity (which gives the first term of (1.1) as a first approximation to EX) and coupling (which gives the second term of (1.1) as a correction for "boundary overshoot error" in the first term). A bound on the difference between EX and (1.1) can be given in terms of, say, $P\{\max_{i\leq X} K_i > \frac{m}{2}\}$ and the probability that a certain Markov-chain coupling is unsuccessful.

Section 4 gives some explicit bounds on the approximation error. These bounds are generally very crude, but they show that the approximation error converges to zero faster than $\exp(-m^{\frac{1}{3}})$ as $m \to \infty$ when the (fixed) K-distribution has a finite moment generating function near 0. If the K_i's are bounded, or if the hazard function of the K_i's is bounded below by $\delta > 0$, then the approximation error converges to zero exponentially (in m) as $m \to \infty$.

Section 5 presents a generalization of (1.1) applicable to the case where some of the balls in the box are red to begin with, and where only a specified number of the white balls need to be painted red.

2. Application of the Wald Identity

Assume the balls are numbered $1, 2, \ldots, m$. Sampling k balls without replacement can of course be done by sampling *with* replacement until k distinct balls have been drawn, with repetitions ignored. This section will relate the sampling-without-replacement scenario described in the introduction to the process of repeatedly drawing balls one at a time, with replacement. Analyzing the number τ of single-ball draws needed to see all the balls is easy: it's just the standard coupon collector problem. Except for "boundary overshoot error," the Wald identity will give an expression for EX in terms of $E\tau$.

Suppose our successive samples of balls are obtained as follows. First we see the value of K_1. Then we sample balls one at a time, with replacement, until $K_1 \wedge m$ distinct balls have been obtained. Let D_1 be the number of single-ball draws needed to obtain the required $K_1 \wedge m$ distinct balls. Then we see K_2 and sample balls one at a time, with replacement, until $K_2 \wedge m$ distinct balls have been obtained to form the second sample, etc. Let D_i be the number of single-ball draws needed to obtain the $K_i \wedge m$ distinct balls in the i^{th} sample. Let $\mathcal{F}_i$ be the σ-field generated by all observations made while generating the first i samples. Thus, $\mathcal{F}_i$ is generated by $K_1, K_2, \ldots, K_i$ *and* by the sequence of $D_1 + \ldots + D_i$ single-ball draws needed to obtain the first i samples. The K_i's are of course assumed to be iid. The ball numbers of the balls obtained on successive draws are iid, uniformly distributed on $\{1, 2, \ldots, m\}$, and independent of the K_i's. The D_i's are also iid.

Let τ be the number of single-ball draws needed to see every ball at least once. The number of additional single-ball draws needed to see a new ball after j balls have already been seen is a geometric $(\frac{m-j}{m})$ random variable independent of everything that has come before. (This is the standard coupon collector argument.) Adding up expectations yields

$$E\tau = \sum_{j=0}^{m-1} \frac{m}{m-j} = m \sum_{i=1}^{m} \frac{1}{i}. \tag{2.1}$$

Again letting X be the number of samples needed to see all m balls, we see that

$$X = \inf\{i : D_1 + D_2 + \ldots + D_i \geq \tau\}.$$

Furthermore, X is an $\mathcal{F}_i$ stopping time. Thus, by Wald's identity and the fact that the D_i's are iid,

$$E(D_1 + \ldots + D_X) = EX\ ED_1. \tag{2.2}$$

The expectation ED_1 is easy to find in terms of the distribution of K_1. The coupon collector argument shows

$$E(D_1 | K_1 \wedge m = k) = \sum_{i=0}^{k-1} \frac{m}{m-i}$$

so

$$\begin{aligned} E(D_1) &= \sum_{k=1}^{m} P\{K_1 \wedge m = k\} \sum_{i=0}^{k-1} \frac{m}{m-i} \\ &= \sum_{i=0}^{m-1} \frac{m}{m-i} \sum_{k=i+1}^{m} P\{K_1 \wedge m = k\} \\ &= \sum_{i=0}^{m-1} \frac{m}{m-i} P\{K > i\}. \end{aligned} \tag{2.3}$$

Now we need to relate $E(D_1 + \ldots + D_X)$ to $E\tau$. The sum $\sum_1^X D_i$ and τ are of course equal except for the "overshoot" given by the number of draws needed to complete the last sample after the last new ball has been obtained on the τ^{th} draw. Let V be the size of this overshoot, so that

$$V = \left(\sum_1^X D_i\right) - \tau. \tag{2.4}$$

From (2.1), (2.2), (2.3), and (2.4) we get

$$\begin{aligned} EX &= \frac{E\sum_1^X D_i}{ED_1} \\ &= \frac{E\tau + EV}{ED_1} \\ &= \frac{m\sum_{i=1}^m \frac{1}{i} + EV}{\sum_{i=0}^{m-1} \frac{m}{m-i} P\{X > i\}}. \end{aligned} \tag{2.5}$$

Let J be the number of distinct balls already in the last sample after the τ^{th} draw. Given that $J = j$ and $K_X = k$ (which is only possible for $k \geq j$), the coupon collector argument shows

$$E(V|J = j, K_X = k) = \sum_{r=j}^{k-1} \frac{m}{m-r} \tag{2.6}$$

(If $k = j$, the right side is supposed to be zero.) Furthermore, the conditional distribution of K_X given that $J = j$ is exactly the same as that of K_1 given that $K_1 \geq j$. (This is perhaps even more obvious when the sampling is done as described in Section 3.) Thus,

$$\begin{aligned} E(V|J = j) &= \sum_{k=j+1}^{m} P\{K = k | K \geq j\} \sum_{r=j}^{k-1} \frac{m}{m-r} \\ &= \sum_{r=j}^{m-1} \frac{m}{m-r} \sum_{k=r+1}^{m} P\{K = k | K \geq j\} \\ &= \sum_{r=j}^{m-1} \frac{m}{m-r} P\{K > r | K \geq j\} \end{aligned} \tag{2.7}$$

(Note that (2.7) is zero for $j = m$.) If we can get a good approximation for the distribution of J, we can combine this with (2.7) to approximate EV in (2.5). The next section will show that $P\{J = j\}$ is typically approximately proportional to

$$\frac{m}{m-(j-1)} P\{K > j-1\} = \frac{m}{m-j+1} P\{K \geq j\},$$

so that

$$\text{(2.8)} \qquad \begin{aligned} P\{J = j\} &\approx \frac{\frac{m}{m-j+1} P\{K \geq j\}}{\sum\limits_{\ell=1}^{m} \frac{m}{m-(\ell-1)} P\{K > \ell - 1\}} \\ &= \frac{\frac{m}{m-j+1} P\{K \geq j\}}{\sum\limits_{i=0}^{m-1} \frac{m}{m-i} P\{K > i\}}. \end{aligned}$$

Combining (2.8) with (2.7) yields

$$\text{(2.9)} \qquad \begin{aligned} EV &= \sum_{j=1}^{m-1} E(V|J = j) P\{J = j\} \\ &\approx \frac{\sum\limits_{j=1}^{m-1} \sum\limits_{r=j}^{m-1} \frac{m}{m-r} P\{K > r | K \geq j\} \frac{m}{m-j+1} P\{K \geq j\}}{\sum\limits_{i=0}^{m-1} \frac{m}{m-i} P\{K > i\}} \\ &= \frac{\sum\limits_{r=1}^{m-1} \frac{m}{m-r} P\{K > r\} \sum\limits_{j=1}^{r} \frac{m}{m-j+1}}{\sum\limits_{i=0}^{m-1} \frac{m}{m-i} P\{K > i\}}. \end{aligned}$$

Substituting this approximation for EV into (2.5) gives (1.1).

3. Approximation of $P\{J = j\}$ by Coupling

This section will justify the approximation (2.8) above:

$$\text{(3.1)} \qquad P\{J = j\} \approx \frac{\frac{m}{m-j+1} P\{K_1 \geq j\}}{\sum\limits_{i=0}^{m-1} \frac{m}{m-i} P\{K > i\}}, \quad j = 1, 2, \ldots, m.$$

The idea is to construct a finite-state-space Markov chain with m absorbing states labelled $1, 2, \ldots, m$ for which J equals the label of the state where absorption occurs. This chain is coupled to an approximating chain for which the distribution of the absorbing state is given (modulo truncation) by the right side of (3.1). (For a general account of the coupling technique, see Lindvall (1992).)

For the purposes of this section, it will be better to use a recipe for generating the required samples which is different from that of the previous section. To start, draw a single ball from the box. Then ask whether the next ball should continue the first sample. The probability of continuing the first sample should be

$$c_1 =: \frac{P\{K \wedge m > 1\}}{P\{K \wedge m \geq 1\}}. \tag{3.2}$$

If we decide to continue the first sample, draw another ball from the box without replacing the first ball. With probability

$$h_1 =: 1 - c_1 = \frac{P\{K \wedge m = 1\}}{P\{K \wedge m \geq 1\}}, \tag{3.3}$$

we decide to end the first sample with the first ball. In this case, we return the first ball to the box and then draw the first ball of the second sample.

The general rule for sampling goes as follows. If the number of balls already in the current sample is a, we decide to continue this sample with the next ball with probability

$$c_a =: \frac{P\{K \wedge m > a\}}{P\{K \wedge m \geq a\}}. \tag{3.4}$$

With probability

$$h_a =: 1 - c_a = \frac{P\{K \wedge m = a\}}{P\{K \wedge m \geq a\}}, \tag{3.5}$$

we end the current sample, return all balls in this current sample to the box, and then draw the first ball of the next sample. (Note that h_a is the hazard function of the K distribution.) Balls are returned to the box only when we decide that the current sample is complete and that it's time to start a new one. It should be obvious that this protocol really does generate independent samples whose common sample size distribution is as required.

Let $\mathcal{G}_n$ be the σ-field generated by everything that happens up to and including the n^{th} ball-draw (but *not* including the decision as to whether the n^{th} ball is the last ball of the sample containing it). Let $\tilde{A}_n$ be the number of balls *already* in the current sample after the n^{th} ball-draw. Let $\tilde{V}_n$ be the number of "virgin" balls left to be drawn for the first time after the n^{th} draw. Define the $\mathcal{G}_n$ stopping time T_0 by

$$T_0 = \inf\{n: \tilde{V}_n = 0\}, \tag{3.6}$$

so that T_0 is the number of ball draws needed to get all balls at least once (according to the sampling protocol of this section.) Note that $\tilde{A}_{T_0} = J$. Define

$$A_n =: \tilde{A}_{n \wedge T_0} \text{ and } V_n =: \tilde{V}_{n \wedge T_0}. \tag{3.7}$$

Then $\{(A_n, V_n)\}_{n=1}^{\infty}$ is a Markov chain adapted to $\mathcal{G}_n$. Since we start with $n = 1$, the starting state is$(1, m-1)$. The state space is

$$S_m =: \{(a, v): a \in \{1, 2, \ldots, m\},\ v \in \{0, 1, \ldots, m-1\},\ a + v \leq m\}.$$

For $v \geq 1$, the transition probabilities are

$$\begin{aligned} P\{(A_{n+1}, V_{n+1}) = (a', v') | (A_n, V_n) = (a, v)\} & \\ = h_a \tfrac{m-v}{m} \quad & \text{for } (a', v') = (1, v) \\ = h_a \tfrac{v}{m} \quad & \text{for } (a', v') = (1, v-1) \\ = c_a \tfrac{m-a-v}{m-a} \quad & \text{for } (a', v') = (a+1, v) \\ = c_a \tfrac{v}{m-a} \quad & \text{for } (a', v') = (a+1, v-1) \\ = 0 \quad & \text{otherwise.} \end{aligned} \tag{3.8}$$

Since the (A_n, V_n) chain stops at time T_0, states of the form $(a, 0)$ are of course absorbing, so that

$$\begin{aligned} P\{(A_{n+1}, V_{n+1}) = (a', v') | (A_n, V_n) = (a, 0)\} & \\ = 1 \ \text{if } (a', v') = (a, 0) & \\ = 0 \ \text{otherwise.} & \end{aligned} \tag{3.9}$$

Since $A_{T_0} = J$, we can analyze the behavior of J by getting the approximate probabilities of absorption at each of the m absorbing states $(a, 0)$.

Fix $b \in \{1, 2, \ldots, m-1\}$. For $a = 1, 2, \ldots, m$, define

$$c_a^{(b)} = \begin{cases} c_a & \text{if } a < b \\ 0 & \text{if } a \geq b \end{cases} \tag{3.10}$$

and

$$h_a^{(b)} = \begin{cases} h_a & \text{if } a < b \\ 1 & \text{if } a \ge b \end{cases} \tag{3.11}$$

Let $(A_n^{(b)}, V_n^{(b)})$ be a Markov chain on S_m whose transition probabilities are given by (3.8) and (3.9), except with $c_a^{(b)}$ and $h_a^{(b)}$ in place of c_a and h_a. Let $\mathcal{G}_n^{(b)}$ be the filtration for this chain.

As far as the transition probabilities are concerned, the $(A_n^{(b)}, V_n^{(b)})$ chain is like the (A_n, V_n) chain with $K \wedge b$ in place of K as the generic sample size. If $P\{K > b\} = 0$, the transition probabilities for the two chains are of course exactly the same. Our coupling construction below will work well when we can choose a b not too close to m for which $P\{K > b\}$ is so small that $P\{\max_{i \le X} K_i > b\}$ is itself negligible.

The $\{(A_n^{(b)}, V_n^{(b)})\}_{n=1}^{\infty}$ Markov chain will be started with $V_1^{(b)} = m - b$. The value of $A_1^{(b)}$ will be random, with

$$P\{A_1^{(b)} = a\} = \frac{\frac{m}{m-a+1} P\{K \ge a\}}{\sum_{i=0}^{b-1} \frac{m}{m-i} P\{K > i\}}, \quad 1 \le a \le b. \tag{3.12}$$

(Compare (3.12) to (3.1). The distribution for $A_1^{(b)}$ in (3.12) is the distribution on the right side of (3.1), conditioned to be $\le b$.)

For $v \in \{0, 1, \ldots, m - b\}$, define

$$T_v^{(b)} = \inf\{n : V_n^{(b)} = v\}. \tag{3.13}$$

Since $V_n^{(b)} - V_{n-1}^{(b)}$ is always either 0 or -1,

$$1 = T_{m-b}^{(b)} > T_{m-b-1}^{(b)} > \ldots > T_1^{(b)} > T_0^{(b)}. \tag{3.14}$$

Here is the key result for our justification of (3.1):

Proposition 3.15.

If $V_1^{(b)} = m - b$ and $A_1^{(b)}$ has the distribution given by (3.12), then $A_{T_v^{(b)}}^{(b)}$ has distribution (3.12) for *all* $v \in \{0, 1, \ldots, m - b\}$. In particular, the first coordinate $A_{T_0^{(b)}}^{(b)}$ of the state $(A_{T_0^{(b)}}^{(b)}, 0)$ where absorption occurs has distribution (3.12).

Proof.

Define $\{B_n^{(b)}\}_{n=1}^{\infty}$ to be a Markov chain on the state space $\{1, 2, \ldots, b\}$ which acts like a non-stopped version of $A_n^{(b)}$. Thus, the transition probabilities for the $B_n^{(b)}$ chain are

$$
\begin{aligned}
(3.16) \qquad P\{B_{n+1}^{(b)} = a' | B_n^{(b)} = a\}& \\
= c_a^{(b)} &= \tfrac{P\{K \wedge b > a\}}{P\{K \wedge b \geq a\}} \text{ if } a' = a+1 \\
= h_a^{(b)} &= \tfrac{P\{K \wedge b = a\}}{P\{K \wedge b \geq a\}} \text{ if } a' = 1 \\
= 0 & \qquad \text{otherwise.}
\end{aligned}
$$

It is well-known (and trivial to check) that the stationary distribution of this chain is

$$
(3.17) \qquad \pi_a = \frac{P\{K \geq a\}}{\sum\limits_{i=1}^{b} P\{K \geq i\}} = \frac{P\{K \geq a\}}{E(K \wedge b)}, \ 1 \leq a \leq b.
$$

Define another Markov chain $(B_n^{(b)}, W_n^{(b)})$ by having $B_n^{(b)}$ as above and the $W_n^{(b)}$'s a sequence of Bernoulli random variables. Set $W_1^{(b)} \equiv 1$. Conditional on the entire path

$$
\underline{B}^{(b)} =: (B_1^{(b)}, B_2^{(b)}, \ldots)
$$

of the $B_n^{(b)}$ chain, the $W_n^{(b)}$'s, $n \geq 2$, are to be independent, with

$$
(3.18) \qquad P\{W_n^{(b)} = 1 | \underline{B}^{(b)}\} = 1 - P\{W_n^{(b)} = 0 | \underline{B}^{(b)}\} = \frac{d}{m - B_n^{(b)} + 1}
$$

for some constant d, $0 < d \leq m - b$.

Now consider the "embedded chain" whose consecutive states are the consecutive states of the $(B_n^{(b)}, W_n^{(b)})$ chain at which $W_n^{(b)} = 1$. (The times at which $W_n^{(b)} = 0$ are skipped.) The stationary distribution for $B_n^{(b)}$ in this embedded chain is

$$
(3.19) \qquad \pi_a^{\mathbf{emb}} = \frac{\frac{1}{m-a+1}\pi_a}{\sum\limits_{i=1}^{u} \frac{1}{m-i+1}\pi_i} = \frac{\frac{1}{m-a+1}P\{K \geq a\}}{\sum\limits_{i=1}^{u} \frac{1}{m-i}P\{K > i\}}, \ 1 \leq a \leq b.
$$

(Note that (3.12) and (3.19) are the same distribution.) The reasoning for (3.19) is as follows. The stationary distribution of an ergodic Markov chain gives the long-run fraction of the time that the chain spends in each state. By (3.17), (3.18), and the strong law of large numbers, the long-run fraction of the time that the $(B_n^{(b)}, W_n^{(b)})$ chain spends in a state $(a, 1)$ equals

$$\pi_a \frac{d}{m-a+1}.$$

Thus, the long-run fraction of the time that the *embedded* chain spends in a state $(a, 1)$ is given by (3.19).

Recall that if the starting state of an ergodic Markov chain is chosen according to the stationary distribution, then the state one time unit later will also have the stationary distribution.

Now compare the $(A_n^{(b)}, V_n^{(b)})$ chain with the $(B_n^{(b)}, W_n^{(b)})$ chain. The differences $V_{n-1}^{(b)} - V_n^{(b)}$ are Bernoulli random variables. For $n \leq T_{m-b-1}^{(b)}$,

$$P\{V_{n-1}^{(b)} - V_n^{(b)} = 1 | \mathcal{G}_{n-1}^{(b)}, A_n^{(b)}\} = \frac{m-b}{m-A_n+1},$$

which looks like (3.18) with $d = m - b$. Thus, up until time

$$T_{m-b-1}^{(b)} = \inf\{n = V_{n-1}^{(b)} - V_n^{(b)} = 1\},$$

$(A_n^{(b)}, V_{n-1}^{(b)} - V_n^{(b)})$ is a chain which acts just like $(B_n^{(b)}, W_n^{(b)})$. Since $A_{T_{m-b-1}^{(b)}}^{(b)}$ corresponds to the second value of $B_n^{(b)}$ in the embedded chain, $A_{T_{m-b-a}^{(b)}}^{(b)}$ will have the stationary distribution (3.12) = (3.19) if $A_1^{(b)}$ starts out in this stationary distribution.

It follows in the same way that $A_{T_{i-1}^{(b)}}^{(b)}$ has distribution (3.12) if $A_{T_i^{(b)}}^{(b)}$ has distribution (3.12). Thus, Proposition 3.15 follows by induction.

□

Now the idea is to show that, with high probability, (A_n, V_n) and $A_n^{(b)}, V_n^{(b)})$ can be coupled so as to end up at the same absorbing state. We start by choosing a starting state for the $(A_n^{(b)}, V_n^{(b)})$ chain as described above, with $V_1^{(b)} = m - b$. Then we let the (A_n, V_n) chain (which always starts in state $(1, m-1)$) run until $V_n = m - b$, without the $(A_n^{(b)}, V_n^{(b)})$ chain moving. Once the chains are on the same V-level, we can sequentially choose one or the other chain to take a step, with the goal of course being to get them to inhabit the same state at the same time. Or, we can let both chains move simultaneously, with the transitions for the two chains being dependent if we want. The requirement is only that, each time a chain takes a step, it must do so according to its own transition probabilities; this will guarantee that the path of each individual chain has the right probabilities. (In particular, the distribution of the absorbing state will be correct.) If we can get the two chains to meet, we couple them so that they move together. Since the transition probabilities are the same as long as A_n and $A_n^{(b)}$ are $< b$, the chains stay coupled as long as the common A-coordinate is $< b$. A coupling can fall apart, however. If the common A-value of the two coupled chains reaches b, then with probability c_b the chains will be decoupled after the next step. If one chain has its V coordinate decrease before coupling is achieved on a common V-level, we just leave this chain alone for a while and run the other chain until its V coordinate also drops, after which we try to achieve a coupling at this next lower V-level.

It does not seem easy to describe (or determine) optimal coupling schemes. However, even very simple-minded strategies can sometimes have very high probabilities of successful coupling.

Let $A_\infty = A_{T_0}$ and $A_\infty^{(b)} = A_{T_0^{(b)}}^{(b)}$ denote the A-values of the absorbing states for the two Markov chains. Let

$$||P\{A_\infty \in \cdot\} - P\{A_\infty^{(b)} \in \cdot\}|| = \sum_{j=1}^{m} |P\{A_\infty = j\} - P\{A_\infty^{(b)} = j\}|$$

denote the total variation distance between the distributions of A_∞ and $A_\infty^{(b)}$.

Proposition 3.20.

If $P\{K > b\} = 0$, then

$$||P\{A_\infty \in \cdot\} - P\{A_\infty^{(b)} \in \cdot\}|| \leq 2\exp(-\frac{m-2b}{b}) < 15e^{-m/b},$$

where the distribution of $A_\infty^{(b)}$ is given by the right side of (3.12).

Proof.

The total variation distance is obviously bounded by twice the probability that the A_n and $A_n^{(b)}$ chains are not coupled when they are absorbed. (cf Lindvall (1992), page 12.)

Here is a coupling strategy which has probability less than $e(-\frac{m-2b}{b})$ that the chains are not coupled at absorption. For the (at most) b possible states in S_m with V coordinate v, order the states as follows:

$$(2,v) \prec (3,v) \prec \ldots \prec (b,v) \prec (1,v).$$

When both chains have their V coordinates equal to v, run the chain which is "behind" according to this ordering and leave the other chain alone. If the chains meet, couple them. (Since $A_n > b$ is impossible, the coupling will not be broken later.) If one chain's V coordinate decreases (by 1), run the other chain until its V coordinate decreases by 1 also.

If no decrease in a V coordinate occurs for the first $b-1$ steps taken when the Markov chains are both on V-level v, then the chains are *guaranteed* to couple on this V-level. Each time a chain takes a step starting on V-level v, the probability that its V coordinate will *not* decrease is at least $\frac{m-b-v}{m-b}$. (See (3.8).) Thus, the probability of *no* coupling on V-level v is at most

$$1-(\frac{m-b-v}{m-b})^{b-1}.$$

The probability that coupling never occurs on *any* of the V-levels $m-b$, $m-b-1,\ldots,2,1$ is at most

$$\prod_{v=1}^{m-b}\{1-(\frac{m-b-v}{m-b})^{b-1}\} = \prod_{i=0}^{m-b-1}\{1-(\frac{i}{m-b})^{b-1}\} \tag{3.21}$$

$$\leq \exp\{-\sum_{i=0}^{m-b-1}(\frac{i}{m-b})^{b-1}\}, \text{ since } 1-x \leq e^{-x}.$$

But

$$\sum_{i=0}^{m-b-1}(\frac{i}{m-b})^{b-1} = \sum_{i=0}^{m-b}(\frac{i}{m-b})^{b-1} - 1 \tag{3.22}$$

$$> (m-b)\int_0^1 x^{b-1}dx - 1 = \frac{m-b}{b} - 1 = \frac{m-2b}{b}.$$

Together, (3.21) and (3.22) imply that $\exp(-\frac{m-2b}{b})$ bounds the probability of the chains not being coupled at absorption.

□

We will see that Proposition 3.20 still holds even when $P\{K_i > b\} > 0$, provided we add $2P\{\max_{i \le X} K_i > b\}$ onto the right side of the inequality in Proposition 3.20. To bound $P\{\max_{i \le X} K_i > b\}$, it will help to have a bound on X.

If we let τ be the number of single-ball draws (with balls replaced after each draw) needed to see every ball at least once (as in Section 2), then it is obviour that $P\{X \ge t\} \le P\{\tau \ge t\}$ for all t. (Recall we assume $P\{K \ge 1\} = 1$.)

Lemma 3.23. $P\{\tau \ge m^2\} < 2m^{\frac{1}{2}}e^{-m/2}$.

Proof.

As was mentioned just before formula (2.1), τ is a sum of independent geometric $(\frac{m-j}{m})$ random variables, $j = 0, 1, \ldots, m-1$. Thus τ has factorial moment generating function

$$\phi_\tau(s) =: E(s^\tau) = \prod_{j=0}^{m-1} \frac{m-j}{m-js} = \prod_{i=1}^{m} \frac{i}{m-(m-i)s}.$$

Setting $s = 1 + \frac{1}{2m}$, we get

$$\begin{aligned}
E(1+\frac{1}{2m})^\tau &= \prod_{i=1}^{m} 1 + \frac{m-i}{2mi-m+i} \\
&< \prod_{i=1}^{m} 1 + \frac{m}{2mi-m} - \frac{i}{2mi} \\
&= \prod_{i=1}^{m} 1 + \frac{1}{2i-1} - \frac{1}{2m} \\
&< \exp\left\{\sum_{i=1}^{m}(\frac{1}{2i-1} - \frac{1}{2m})\right\} \\
&= \exp(-\frac{1}{2} + \sum_{i=1}^{m} \frac{1}{2i-1}).
\end{aligned}$$

But

$$\begin{aligned}\sum_{i=1}^{m} \frac{1}{2i-1} &= 1 + \frac{1}{3} + \frac{1}{5} + \ldots + \frac{1}{2m-1} \\ &< \frac{4}{3} + \frac{1}{2}\int_4^{2m} \frac{1}{x} dx \\ &< 1 + \frac{1}{2} \log\ m.\end{aligned}$$

Thus,

$$E(1 + \frac{1}{2m})^{\tau} < (me)^{\frac{1}{2}}.$$

However,

$$(1 + \frac{1}{2m})^{m^2} > (e^{\frac{1}{2m} - \frac{1}{8m^2}})^{m^2} = e^{\frac{m}{2} - \frac{1}{8}}$$

$$\text{(since } \log\ (1+x) > x - \frac{x^2}{2} \text{ for } 0 \le x \le 1).$$

Thus, by the Markov inequality,

$$P\{\tau \ge m^2\} < (me)^{\frac{1}{2}} e^{\frac{1}{8} - \frac{m}{2}} < 2m^{\frac{1}{2}} e^{-\frac{m}{2}}.$$

□

Corollary 3.24. $P\{X \ge m^2\} < 2m^{\frac{1}{2}} e^{-m/2}$.

Proposition 3.25.

If $P\{K > b\} < \varepsilon$, then

$$||P\{A_\infty \in \cdot\} - P\{A_\infty^{(b)} \in \cdot\}|| < 15e^{-m/b} + 2m^2\varepsilon + 4m^{\frac{1}{2}} e^{-m/2}.$$

Proof.

Use the same coupling strategy as in the proof of Proposition 3.20. Expression (3.21) (and therefore $\frac{15}{2} e^{-m/b}$) bounds the probability of the event "no (initial) coupling occurs *and* $\max\limits_{i \le X} K_i \le b$." But by Corollary 3.24,

$$(3.26) \qquad \begin{aligned} P\{\max_{i \ge X} K_i > b\} &\le m^2 P\{K > b\} + P\{X > m^2\} \\ &< m^2 \varepsilon + 3m^{\frac{1}{2}} e^{-m/2}.\end{aligned}$$

Thus the probability of "no (initial) coupling occurs *or* $\max_{i \le X} K_i > b$" is bounded by

$$\frac{15}{2}e^{-m/b} + m^2\varepsilon + 2m^{\frac{1}{2}}e^{-m/2}. \tag{3.27}$$

Note that a coupling, once made, is never broken on $\{\max_{i \le X} K_i \le b\}$. Thus, (3.27) bounds the probability that the (A_n, V_n) and $(A_n^{(b)}, V_n^{(b)})$ chains are not coupled at absorption. Multiplying (3.27) by 2 gives the desired bound on the total variation distance.

□

If the hazard function h_a of the K_i's is bounded below, a different coupling rule sometimes works better than the one used above. The bound in the next proposition should be particularly good if the hazard function h_a is (approximately) constant, so that the K_i's are (approximately) geometric random variables.

Proposition 3.28.

If the hazard function h_a of the K_i's is bounded below by $\delta > 0$ for $a < b$, and if $(m-b)\delta \ge 1$, then

$$||P\{A_\infty \in \cdot\} - P\{A_\infty^{(b)} \in \cdot\}|| \le 6\sqrt{m-b}\left\{\frac{\delta^\delta}{(1+\delta)^{1+\delta}}\right\}^{m-b} + 4m^{\frac{1}{2}}e^{-m/2} + 2m^2(1-\delta)^b.$$

Remark. If the K_i's are geometric $(\frac{1}{2})$ and $m = 100$, then taking $b = 62$ and $\delta = \frac{1}{2}$ causes the bound in Proposition 3.28 to equal 1.08 $\times 10^{-14}$.

Proof.

Again, if the Markov chains are on different V-levels, run only the one on the higher V-level until it drops down to the V-level where the other is. When both chains are on the same V-level, the strategy here will be to have the chains move simultaneously and dependently. The goal will be to get both chains to jump to state $(1, v)$ at the same time, or to jump to state $(1, v-1)$ at the same time.

Lemma 3.29. Suppose the (A_n, V_n) and $(A_n^{(b)}, V_n^{(b)})$ chains are both on V-level v in different states. Then by having the chains take dependent, simultaneous steps, it is possible to make the (conditional, given the past) probability of the event "A_n does not exceed b *and* the chains don't couple on or before the first time that at least one chain drops to V-level $v-1$" less than or equal to $\frac{v}{v+(m-b)\delta}$.

Proof of Lemma 3.29.

We need to describe the dependence between the next step of the (A_n, V_n) chain and the next step of the $(A_n^{(b)}, V_n^{(b)})$ chain. We first determine the next A values. By assumption, when $A_n < b$ each chain has probability $\geq \delta$ of having its A coordinate jump to 1 on the next step. If $A_n = b$, then A_{n+1} will either equal 1 or $b+1$, and the $(A_n^{(b)}, V_n^{(b)})$ chain of course still has probability $\geq \delta$ of the next A value being 1.

Thus, when $A_n \leq b$ we can choose the next A values in such a way that the event "either $A_{n+1} > b$, or both chains have their next A values equal to 1" has probability $\geq \delta$. After determining the new A values we determine the new V values. If both new A values are 1, then for both chains the conditional probability is $\frac{m-v}{m}$ that the new V value is v and $\frac{v}{m}$ that the new V value is $v-1$. Thus, when both new A values are 1, we can choose the new V values to be equal as well, so the chains couple. Providing that the new A value for the (A_n, V_n) chain is $\leq b$, the conditional probability (given its new A value) for each individual chain to drop to V-level $v-1$ is $\leq \frac{v}{m-b}$. Thus, we can make the new V values dependent in such a way that the probability of a V-level decrease in either chain is $\leq \frac{v}{m-b}$.(unless $A_{n+1} > b$, in which case we don't care about V values.)

If the consecutive steps of the two chains are made dependent as described in the previous paragraph, then the event "the A_n's do not exceed b *and* the chains do not couple on or before the first time a chain leaves V-level v" has probability less than or equal to

$$\frac{\frac{v}{m-b}}{\frac{v}{m-b}+\delta} = \frac{v}{v+(m-b)\delta}.$$

□

Continuation of Proof of Proposition 3.28.

Lemma 3.29 immediately implies that the event "the chains never couple (on *any*

V-level) and A_n never exceeds b (on *any* V-level)" has probability less than or equal to

$$\prod_{v=1}^{m-b} \frac{v}{v+(m-b)\delta} = \frac{(m-b)!\Gamma\{(m-b)\delta+1\}}{\Gamma\{(m-b)(1+\delta)+1\}} \tag{3.30}$$

where $\Gamma(\cdot)$ is the gamma function. To estimate (3.30), we can use

$$\Gamma(x+1) = \sqrt{2\pi}\; x^{x+\frac{1}{2}} e^{-x} e^{\theta(x)/12x}, \; x > 0, 0 \leq \theta(x) \leq 1.$$

(See the *Encyclopedia of Statistical Sciences*, (1988), Wiley, Kotz and Johnson, Eds., vol. 8, p. 779.) Applying this form of Stirling's formula to (3.30) yields

$$\begin{aligned}\prod_{v=1}^{m-b} \frac{v}{v+(m-b)\delta} &< \sqrt{2\pi(m-b)} \left\{ \frac{\delta^\delta}{(1+\delta)^{1+\delta}} \right\}^{m-b} \exp\{\frac{1}{6(m-b)\delta}\} \\ &< 3\sqrt{m-b} \left\{ \frac{\delta^\delta}{(1+\delta)^{1+\delta}} \right\}^{m-b} \text{ if } (m-b)\delta \geq 1.\end{aligned}$$

Since an achieved coupling is never broken on $\{\max_{i \leq X} K_i \leq b\}$, the probability of the chains not being coupled at absorption is bounded by

$$3\sqrt{m-b} \left\{ \frac{\delta^\delta}{(1+\delta)^{1+\delta}} \right\}^{m-b} + P\{\max_{i \leq X} K_i > b\}.$$

Multiplying by 2 and applying Corollary 3.24 (as in (3.26)) produces the bound in Proposition 3.28.

□

4. Bounds on the Approximation Error

Using the bounds on the total variation distance $||P\{A_\infty \in \cdot\} - P\{A_\infty^{(b)} \in \cdot\}||$ given in Section 3, it is easy to derive bounds on the difference between (1.1) and EX. Recall again that J equals the random variable A_∞ from Section 3.

Lemma 4.1. For any $b \in \{1, 2, \ldots, m-1\}$,

$$\begin{aligned}&|EV - \sum_{j=1}^{m} E(V|J=j) P\{A_\infty^{(b)} = j\}| \\ &\qquad \leq \frac{1}{2} ||P\{A_\infty \in \cdot\} - P\{A_\infty^{(b)} \in \cdot\}|| \max_j E(V|J=j),\end{aligned}$$

where $E(V|J=j)$ is set equal to 0 when $P\{K \geq j\} = 0$.

Proof.

The left side equals

$$|\sum_{j=1}^{m} E(V|J=j)[P\{A_\infty = j\} - P\{A_\infty^{(b)} = j\}]|,$$

which is less than or equal to the right side.

□

Lemma 4.2. For any $b \in \{1, \ldots, m-1\}$

$$\left| \frac{\sum\limits_{r=1}^{m-1} \frac{m}{m-r} P\{K > r\} \sum\limits_{j=1}^{r} \frac{m}{m-j+1}}{\sum\limits_{i=0}^{m-1} \frac{m}{m-i} P\{K > i\}} - \frac{\sum\limits_{r=1}^{b-1} \frac{m}{m-r} P\{K > r\} \sum\limits_{j=1}^{r} \frac{m}{m-j+1}}{\sum\limits_{i=0}^{b-1} \frac{m}{m-i} P\{K > i\}} \right|$$

$$\leq \frac{\sum\limits_{r=b}^{m-1} \frac{m}{m-r} P\{K > r\} \sum\limits_{j=1}^{r} \frac{m}{m-j+1}}{\sum\limits_{i=0}^{b-1} \frac{m}{m-i} P\{K > i\}}$$

Remark. The first term in Lemma 4.2 is the approximation (2.9) for EV that we used to get (1.1). The second term in Lemma 4.2 equals $\sum\limits_{j=1}^{m} E(V|J=j)P\{A_\infty^{(b)} = j\}$.

Proof.

$\sum\limits_{j=1}^{r} \frac{m}{m-j+1}$ is obviously increasing in r, so the first term between absolute value signs is larger than the second term. The first numerator is greater than the second numerator, and the first denominator is greater than the second denominator. Thus, the difference is less than the difference between numerators divided by the second (smaller) denominator.

□

Proposition 4.3.

The difference between EX and (1.1) is bounded in absolute value by

$$\min_{0<b<m}\left\{\frac{1}{2}\|P\{A_\infty\in\cdot\}-P\{A_\infty^{(b)}\in\cdot\}\|\max_{j\geq 1}\left[\sum_{r=j}^{m-1}\frac{m}{m-r}P\{K>r|K\geq j\}\right]\right.$$
$$\left.+\frac{\sum_{r=b}^{m-1}\frac{m}{m-r}P\{K>r\}\sum_{j=1}^{r}\frac{m}{m-j+1}}{\sum_{i=0}^{b-1}\frac{m}{m-i}P\{K>i\}}\right\}\Big/\sum_{i=0}^{m-1}\frac{m}{m-i}P\{K>i\}.$$

Proof.

This bound follows from (2.5) and Lemmas 4.1 and 4.2.

□

Now let's specialize to get bounds which are not quite such a horrendous mess like the bound in Proposition 4.3.

Proposition 4.4.

If $P\{K>b\}=0$, the difference between EX and (1.1) is bounded in absolute value by

$$\frac{\frac{15}{2}e^{-m/b}\frac{m(b-1)}{m-b}}{\sum_{i=0}^{b-1}\frac{m}{m-i}P\{K>i\}}<\frac{15m(b-1)e^{-m/b}}{2(m-b)EK}.$$

Proof.

If $P\{K>b\}=0$, then

$$\max_{j\geq 1}\sum_{r=j}^{m-1}\frac{m}{m-r}P\{K>r|K\geq j\}\leq\frac{m(b-1)}{m-b}.$$

Apply this and Proposition 3.20 to the bound in Proposition 4.3.

□

Remark. If $m = 100$ and each $K_i - 1$ is binomial $(4, \frac{1}{2})$, the second error bound in Proposition 4.4 equals 1.96×10^{-4} (with $b = 5$ and $EK = 3$).

Proposition 4.5.

If $P\{K > k\} \leq Ce^{-\alpha k}$ for $k \leq b$, $C > 0$, and $\alpha > 0$, then the difference between (1.1) and EX is bounded in absolute value by

$$(\frac{15}{2}e^{-m/b} + m^2Ce^{-\alpha b} + 2m^{\frac{1}{2}}e^{-m/2})\frac{m\log m}{E(K \wedge m)} + Ce^{-\alpha b}(\frac{m\log m}{E(K \wedge b)})^2.$$

Remark. If $P\{K > k\} \leq Ce^{-\alpha k}$ for all k, then letting $b \approx \sqrt{m}$ in the Proposition 4.5 bound shows that the approximation error converges to zero faster than $\exp(-m^{1/3})$ as $m \rightarrow \infty$.

Proof.

$$\max_{j \geq 1} \sum_{r=j}^{m-1} \frac{m}{m-r} P\{K > r | K \geq j\} \leq m \sum_{r=1}^{m-1} \frac{1}{m-r} < m\log m.$$

Also,

$$\sum_{r=b}^{m-1} \frac{m}{m-r} P\{K > r\} \sum_{j=1}^{r} \frac{m}{m-j+1} \leq Ce^{-\alpha b}(m\log m)^2,$$

$$\sum_{i=0}^{b-1} \frac{m}{m-i} P\{K > i\} \geq E(K \wedge b).$$

and

$$\sum_{i=0}^{m-1} \frac{m}{m-i} P\{K > i\} \geq E(K \wedge m).$$

Combining these bounds and Proposition 3.25 with Proposition 4.3 proves Proposition 4.5.

□

Remark. If $m = 800$ and the K_i's are geometric $(\frac{1}{2})$, then taking $b = 40$ in Proposition 4.5 yields the error bound 0.0016. The next proposition is much more effective for geometric K_i's.

Proposition 4.6.

If the hazard function of the distribution of the K_i's

$$h_a = \frac{P\{K = a\}}{\{K \geq a\}}(=: 1 \text{ for } \frac{0}{0})$$

satisfies $h_a \geq \delta > 0$ for $a \leq m$, then the difference between (1.1) and EX is bounded in absolute value by

$$\left[3(m-b)^{\frac{1}{2}}\left\{\frac{\delta^\delta}{(1+\delta)^{1+\delta}}\right\}^{m-b} + 2m^{\frac{1}{2}}e^{-m/2} + m^2(1-\delta)^b\right]\frac{m(1-\delta)}{\delta} + (1-\delta)^b\left(\frac{m\log\, m}{E(K\wedge b)}\right)^2.$$

Proof.

$h_a \geq \delta \;\forall\; a \leq m$ implies $P\{K > r|K \geq j\} \leq (1-\delta)^{r-j+1}$. Thus,

$$\max_{j\geq 1}\sum_{r=j}^{m-1}\frac{m}{m-r}P\{K > r|K \geq j\} \leq \max_{j\geq 1} m\sum_{r=j}^{m-1}(1-\delta)^{r-j+1} < \frac{m(1-\delta)}{\delta}.$$

Combining this with Proposition 3.28 gives us the first term above as a bound on the first term in Proposition 4.3. The second term in Proposition 4.6 follows in the same way as the second term in Proposition 4.5.

□

Remark. If $m = 100$ and the K_i's are geometric $(\frac{1}{2})$, then taking $b = 62$ and $\delta = \frac{1}{2}$ causes the bound in Proposition 4.6 to be 5.64×10^{-13}.

5. Generalization

Suppose we start with w white balls and $m - w$ red balls in the box. Let $Y_{w,n}$ be the number of iid samples needed to see (or paint red) n of the white balls. (Thus, $X = Y_{m,m}$.) The formula

$$(5.1) \qquad \frac{\sum_{i=w-n+1}^{w}\frac{1}{i}}{\sum_{i=0}^{m-1}\frac{1}{m-i}P\{K > i\}} + \frac{\sum_{r=1}^{m-1}\frac{1}{m-r}P\{K > r\}\sum_{j=1}^{r}\frac{1}{m-j+1}}{\left[\sum_{i=0}^{m-1}\frac{1}{m-i}P\{K > i\}\right]^2}$$

should (usually) be a good approximation for $EY_{w,n}$ when the first term of (5.1) is large. The argument for the first term is exactly the same as in Section 2: the numerator is the expected number of single-ball draws needed to get the required number of white balls, and the denominator is the expected number of draws needed to complete a sample. The second term of (5.1) (which is exactly the same as the second term of (1.1)) is again a correction for the error in the first term caused by "boundary overshoot." In this case, the $\{(A_n, V_n)\}_{n=1}^{\infty}$ chain of Section 3 starts either in state $(1, w)$ or in state $(1, w-1)$, depending on whether the first ball chosen is red or white. The (A_n, V_n) chain is absorbed by the states on V-level $w-n$. If a successful coupling can be achieved with high probability between this (A_n, V_n) chain and an $(A_n^{(b)}, V_n^{(b)})$ chain, then the A coordinate of the absorbing state will have a distribution approximated by (3.12). Providing also that the bound in Lemma 4.2 is small, the second term of (5.1) will do a good job of correcting the error in the first term.

References

Kotz, S. and Johnson, N. L., Eds. (1988), *Encyclopedia of Statistical Sciences*, volume 8, New York: John Wiley.

Lindvall, T. (1992), *Lectures on the Coupling Method*, New York: John Wiley.

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 460	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) HOW MANY IID SAMPLES DOES IT TAKE TO SEE ALL THE BALLS IN A BOX?		5. TYPE OF REPORT & PERIOD COVERED Technical 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Thomas M. Sellke		8. CONTRACT OR GRANT NUMBER(s) N0025-92-J-1264
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics Stanford University Stanford, CA 94305-4065		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR-042-267
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Statistics & Probability Program Code 111		12. REPORT DATE October 26, 1992 13. NUMBER OF PAGES 25
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

THE VIEW, OPINIONS, AND/OR FINDINGS CONTAINED IN THIS REPORT ARE THOSE OF THE AUTHOR(S) AND SHOULD NOT BE CONSTRUED AS AN OFFICIAL DEPARTMENT OF THE ARMY POSITION, POLICY, OR DECISION, UNLESS SO DESIGNATED BY OTHER DOCUMENTATION.

THIS REPORT ALSO ISSUED AS TECHNICAL REPT #92-47, Purdue University, 10/92.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

See Reverse Side

Abstract

Suppose a box contains m balls, numbered from 1 to m. A random number of balls are drawn from the box, their numbers are noted, and the balls are then returned to the box. This is done repeatedly, with the sample sizes being iid. Let X be the number of samples needed to see all the balls. This paper derives a simple but typically very accurate approximation for EX in terms of the sample size distribution. The justification of the approximation formula uses Wald's identity and Markov-chain coupling.